

## RAW SEQUENCE LISTING

**The Biotechnology Systems Branch of the Scientific and Technical Information Center (STIC) no errors detected.**

Application Serial Number: 10/810,277  
Source: IFWO  
Date Processed by STIC: 10-18-04

***ENTERED***



IFWO

## RAW SEQUENCE LISTING

DATE: 10/18/2004

PATENT APPLICATION: US/10/810,277

TIME: 10:00:39

Input Set : A:\GC794-2-seqlist.txt

Output Set: N:\CRF4\10182004\J810277.raw

```

4 <110> APPLICANT: Larenas, Edmund A.
5      Goedegebuur, Frits
6      Gualfetti, Peter
7      Mitchinson, Colin
9 <120> TITLE OF INVENTION: Variant Humicola grisea CBH1.1
12 <130> FILE REFERENCE: GC794-2
14 <140> CURRENT APPLICATION NUMBER: US 10/810,277
15 <141> CURRENT FILING DATE: 2004-03-26
17 <150> PRIOR APPLICATION NUMBER: US 60/459,734
18 <151> PRIOR FILING DATE: 2003-04-01
20 <160> NUMBER OF SEQ ID NOS: 15
22 <170> SOFTWARE: FastSEQ for Windows Version 4.0
24 <210> SEQ ID NO: 1
25 <211> LENGTH: 1638
26 <212> TYPE: DNA
27 <213> ORGANISM: Humicola grisea
29 <400> SEQUENCE: 1
30 atgcgtaccg ccaagttcgc caccctcgcc gcccttgtgg cctcggccgc cgcccagcag      60
31 gcgtgcagtc tcaccaccga gaggcaccct tccctctctt ggaagaagtg caccgcccgc      120
32 ggccagtgcc agaccgtcca ggcttccatc actctcgact ccaactggcg ctggactcac      180
33 caggtgtctg gctccacca ctgctacacg ggcaacaagt gggatactag catctgcact      240
34 gatgccaagt cgtgcgctca gaactgctgc gtcgatggcg ccgactacac cagcacctat      300
35 ggcataccca ccaacgggtga ttccctgagc ctcaagttcg tcaccaaggg ccagcactcg      360
36 accaacgtcg gctcgcgtac ctacctgatg gacggcgagg acaagtatca gagtacgttc      420
37 tatcttcagc cttctcgcgc cttgaatcct ggctaacggt tacacttcac agccttcgag      480
38 ctctcggca acgagttcac cttcgatgtc gatgtctcca acatcggtg cggtctcaac      540
39 ggcgcctgt acttcgtctc catggacgcc gatggtggcg tcagccgcta tcctggcaac      600
40 aaggctggcg ccaagtacgg taccggctac tgcgatgtc agtgcccccg tgacatcaag      660
41 ttcatacaac gcgaggccaa cattgagggc tggaccggct ccaccaacga cccaacgcc      720
42 ggcgcggggc gctatggtac ctgctgctct gagatggata tctgggaagc caacaacatg      780
43 gctactgcct tcaactcctc cccttgacc atcattggcc agagccgctg cgagggcgac      840
44 tcgtgcggcg gcacctacag caacgagcgc tacgccggcg tctgcgaccc cgatggctgc      900
45 gacttcaact cgtaccgcca gggcaacaag accttctacg gcaagggcat gaccgtcgac      960
46 accaccaaga agatcactgt cgtcaccag ttctcaagg atgccaacgg cgatctcggc      1020
47 gagatcaagc gcttctacgt ccaggatggc aagatcatcc ccaactccga gtccaccatc      1080
48 cccggcgctc agggcaattc catcaccag gactggcgcg accgccagaa gggtgccttt      1140
49 ggcgacattg acgacttcaa ccgcaagggc ggcatagaag agatgggcaa ggccctcgcc      1200
50 ggcccatagg tcctgggcat gtccatctgg gatgaccacg cctccaacat gctctggctc      1260
51 gactcgacct tcctgtcga tgccgctggc aagcccgcg ccgagcgcg tgctgcccg      1320
52 accacctcgg gtgtccctgc tgaggttgag gccgaggccc ccaacagcaa cgtcgtcttc      1380
53 tccaacatcc gcttcggccc catcggtcgc accgttgctg gtctccccgg cgcgggcaac      1440
54 ggcggcaaca acggcggaac cccccgccc cccaccacca ccactctc ggctccggcc      1500
55 accaccacca ccgccagcgc tggccccaag gctggccgct ggcagcagtg cggcggcac      1560

```

## RAW SEQUENCE LISTING

PATENT APPLICATION: US/10/810,277

DATE: 10/18/2004

TIME: 10:00:39

Input Set : A:\GC794-2-seqlist.txt

Output Set: N:\CRF4\10182004\J810277.raw

```

56 ggctttcactg gcccgaccca gtgcgaggag ccctacactt gcaccaagct caacgactgg      1620
57 tactctcagt gcctgtaa                                          1638
59 <210> SEQ ID NO: 2
60 <211> LENGTH: 1323
61 <212> TYPE: DNA
62 <213> ORGANISM: Humicola grisea
64 <400> SEQUENCE: 2
65 gctcagaact gctgcgtcga tgggtgccgac tacaccagca cctatggcat caccaccaac      60
66 ggtgattccc tgagcctcaa gttcgtcacc aagggccagc actcgaccaa cgtcggctcg      120
67 cgtacctacc tgatggacgg cgaggacaag tatcagacct tcgagctcct cggcaacgag      180
68 ttcaccttcg atgtcgatgt ctccaacatc ggctgcggtc tcaacggcgc cctgtacttc      240
69 gtctccatgg acgccgatgg tgggtctcagc cgctatcctg gcaacaaggc tggtgccaag      300
70 tacggtaccg gctactgcga tgctcagtgc ccccgtgaca tcaagttcat caacggcgag      360
71 gccaacattg agggctggac cgggtccacc aacgaccca acgccggcgc gggccgctat      420
72 ggtacctgct gctctgagat ggatatctgg gaagccaaca acatggctac tgccttcact      480
73 cctcaccctt gcaccatcat tggccagagc cgctgcgagg gcgactcgtg cggtggcacc      540
74 tacagcaacg agcgctacgc cggcgtctgc gaccccgatg gctgcgactt caactcgtac      600
75 cgccaggggca acaagacctt ctacggcaag ggcattgaccg tcgacaccac caagaagatc      660
76 actgtcgtca cccagttcct caaggatgcc aacggcgatc tcggcgagat caagcgcttc      720
77 tacgtccagg atggcaagat catccccaac tccgagtcca ccatccccgg cgtcgagggc      780
78 aattccatca cccaggactg gtgcgaccgc cagaaggttg cttttggcga cattgacgac      840
79 ttcaaccgca agggcggcat gaagcagatg ggcaaggccc tcgccggccc catggtcctg      900
80 gtcattgtcca tctgggatga ccacgcctcc aacatgctct ggctcgactc gaccttcctt      960
81 gtcgatgccg ctggcaagcc cggcgccgag cgcggtgcct gcccgaccac ctcggtgtgc      1020
82 cctgctgagg ttgaggccga ggcccccaac agcaacgtcg tcttctccaa catccgcttc      1080
83 ggccccatcg gctcgaccgt tgctggtctc cccggcgcgg gcaacggcgg caacaacggc      1140
84 ggcaaccccc cgccccccac caccaccacc tcctcggtc cggccaccac caccaccgcc      1200
85 agcgctggcc ccaaggctgg ccgctggcag cagtgcggcg gcatcggtt cactggcccg      1260
86 acccagtgcg aggagcccta cacttgccac aagctcaacg actggtactc tcagtgcctg      1320
87 taa                                          1323
89 <210> SEQ ID NO: 3
90 <211> LENGTH: 525
91 <212> TYPE: PRT
92 <213> ORGANISM: Humicola grisea
94 <400> SEQUENCE: 3
95 Met Arg Thr Ala Lys Phe Ala Thr Leu Ala Ala Leu Val Ala Ser Ala
96 1          5          10          15
97 Ala Ala Gln Gln Ala Cys Ser Leu Thr Thr Glu Arg His Pro Ser Leu
98          20          25          30
99 Ser Trp Lys Lys Cys Thr Ala Gly Gly Gln Cys Gln Thr Val Gln Ala
100          35          40          45
101 Ser Ile Thr Leu Asp Ser Asn Trp Arg Trp Thr His Gln Val Ser Gly
102          50          55          60
103 Ser Thr Asn Cys Tyr Thr Gly Asn Lys Trp Asp Thr Ser Ile Cys Thr
104 65          70          75          80
105 Asp Ala Lys Ser Cys Ala Gln Asn Cys Cys Val Asp Gly Ala Asp Tyr
106          85          90          95
107 Thr Ser Thr Tyr Gly Ile Thr Thr Asn Gly Asp Ser Leu Ser Leu Lys
108          100          105          110

```

## RAW SEQUENCE LISTING

PATENT APPLICATION: US/10/810,277

DATE: 10/18/2004

TIME: 10:00:39

Input Set : A:\GC794-2-seqlist.txt

Output Set: N:\CRF4\10182004\J810277.raw

109	Phe	Val	Thr	Lys	Gly	Gln	His	Ser	Thr	Asn	Val	Gly	Ser	Arg	Thr	Tyr
110			115					120					125			
111	Leu	Met	Asp	Gly	Glu	Asp	Lys	Tyr	Gln	Thr	Phe	Glu	Leu	Leu	Gly	Asn
112		130					135					140				
113	Glu	Phe	Thr	Phe	Asp	Val	Asp	Val	Ser	Asn	Ile	Gly	Cys	Gly	Leu	Asn
114	145					150					155					160
115	Gly	Ala	Leu	Tyr	Phe	Val	Ser	Met	Asp	Ala	Asp	Gly	Gly	Leu	Ser	Arg
116					165					170						175
117	Tyr	Pro	Gly	Asn	Lys	Ala	Gly	Ala	Lys	Tyr	Gly	Thr	Gly	Tyr	Cys	Asp
118				180						185					190	
119	Ala	Gln	Cys	Pro	Arg	Asp	Ile	Lys	Phe	Ile	Asn	Gly	Glu	Ala	Asn	Ile
120			195					200					205			
121	Glu	Gly	Trp	Thr	Gly	Ser	Thr	Asn	Asp	Pro	Asn	Ala	Gly	Ala	Gly	Arg
122		210						215				220				
123	Tyr	Gly	Thr	Cys	Cys	Ser	Glu	Met	Asp	Ile	Trp	Glu	Ala	Asn	Asn	Met
124	225					230					235					240
125	Ala	Thr	Ala	Phe	Thr	Pro	His	Pro	Cys	Thr	Ile	Ile	Gly	Gln	Ser	Arg
126					245					250						255
127	Cys	Glu	Gly	Asp	Ser	Cys	Gly	Gly	Thr	Tyr	Ser	Asn	Glu	Arg	Tyr	Ala
128				260					265						270	
129	Gly	Val	Cys	Asp	Pro	Asp	Gly	Cys	Asp	Phe	Asn	Ser	Tyr	Arg	Gln	Gly
130			275					280					285			
131	Asn	Lys	Thr	Phe	Tyr	Gly	Lys	Gly	Met	Thr	Val	Asp	Thr	Thr	Lys	Lys
132		290					295					300				
133	Ile	Thr	Val	Val	Thr	Gln	Phe	Leu	Lys	Asp	Ala	Asn	Gly	Asp	Leu	Gly
134	305					310					315					320
135	Glu	Ile	Lys	Arg	Phe	Tyr	Val	Gln	Asp	Gly	Lys	Ile	Ile	Pro	Asn	Ser
136				325						330					335	
137	Glu	Ser	Thr	Ile	Pro	Gly	Val	Glu	Gly	Asn	Ser	Ile	Thr	Gln	Asp	Trp
138				340						345					350	
139	Cys	Asp	Arg	Gln	Lys	Val	Ala	Phe	Gly	Asp	Ile	Asp	Asp	Phe	Asn	Arg
140			355					360					365			
141	Lys	Gly	Gly	Met	Lys	Gln	Met	Gly	Lys	Ala	Leu	Ala	Gly	Pro	Met	Val
142		370					375						380			
143	Leu	Val	Met	Ser	Ile	Trp	Asp	Asp	His	Ala	Ser	Asn	Met	Leu	Trp	Leu
144	385					390					395					400
145	Asp	Ser	Thr	Phe	Pro	Val	Asp	Ala	Ala	Gly	Lys	Pro	Gly	Ala	Glu	Arg
146					405					410						415
147	Gly	Ala	Cys	Pro	Thr	Thr	Ser	Gly	Val	Pro	Ala	Glu	Val	Glu	Ala	Glu
148				420						425					430	
149	Ala	Pro	Asn	Ser	Asn	Val	Val	Phe	Ser	Asn	Ile	Arg	Phe	Gly	Pro	Ile
150			435					440						445		
151	Gly	Ser	Thr	Val	Ala	Gly	Leu	Pro	Gly	Ala	Gly	Asn	Gly	Gly	Asn	Asn
152		450					455					460				
153	Gly	Gly	Asn	Pro	Pro	Pro	Pro	Thr	Thr	Thr	Thr	Ser	Ser	Ala	Pro	Ala
154	465					470					475					480
155	Thr	Thr	Thr	Thr	Ala	Ser	Ala	Gly	Pro	Lys	Ala	Gly	Arg	Trp	Gln	Gln
156					485					490					495	
157	Cys	Gly	Gly	Ile	Gly	Phe	Thr	Gly	Pro	Thr	Gln	Cys	Glu	Glu	Pro	Tyr

## RAW SEQUENCE LISTING

PATENT APPLICATION: US/10/810,277

DATE: 10/18/2004

TIME: 10:00:39

Input Set : A:\GC794-2-seqlist.txt

Output Set: N:\CRF4\10182004\J810277.raw

```

158          500          505          510
159 Thr Cys Thr Lys Leu Asn Asp Trp Tyr Ser Gln Cys Leu
160          515          520          525
162 <210> SEQ ID NO: 4
163 <211> LENGTH: 507
164 <212> TYPE: PRT
165 <213> ORGANISM: Humicola grisea
167 <400> SEQUENCE: 4
168 Gln Gln Ala Cys Ser Leu Thr Thr Glu Arg His Pro Ser Leu Ser Trp
169 1          5          10          15
170 Lys Lys Cys Thr Ala Gly Gly Gln Cys Gln Thr Val Gln Ala Ser Ile
171          20          25          30
172 Thr Leu Asp Ser Asn Trp Arg Trp Thr His Gln Val Ser Gly Ser Thr
173          35          40          45
174 Asn Cys Tyr Thr Gly Asn Lys Trp Asp Thr Ser Ile Cys Thr Asp Ala
175          50          55          60
176 Lys Ser Cys Ala Gln Asn Cys Cys Val Asp Gly Ala Asp Tyr Thr Ser
177 65          70          75          80
178 Thr Tyr Gly Ile Thr Thr Asn Gly Asp Ser Leu Ser Leu Lys Phe Val
179          85          90          95
180 Thr Lys Gly Gln His Ser Thr Asn Val Gly Ser Arg Thr Tyr Leu Met
181          100         105         110
182 Asp Gly Glu Asp Lys Tyr Gln Thr Phe Glu Leu Leu Gly Asn Glu Phe
183          115         120         125
184 Thr Phe Asp Val Asp Val Ser Asn Ile Gly Cys Gly Leu Asn Gly Ala
185          130         135         140
186 Leu Tyr Phe Val Ser Met Asp Ala Asp Gly Gly Leu Ser Arg Tyr Pro
187 145         150         155         160
188 Gly Asn Lys Ala Gly Ala Lys Tyr Gly Thr Gly Tyr Cys Asp Ala Gln
189          165         170         175
190 Cys Pro Arg Asp Ile Lys Phe Ile Asn Gly Glu Ala Asn Ile Glu Gly
191          180         185         190
192 Trp Thr Gly Ser Thr Asn Asp Pro Asn Ala Gly Ala Gly Arg Tyr Gly
193          195         200         205
194 Thr Cys Cys Ser Glu Met Asp Ile Trp Glu Ala Asn Asn Met Ala Thr
195          210         215         220
196 Ala Phe Thr Pro His Pro Cys Thr Ile Ile Gly Gln Ser Arg Cys Glu
197 225         230         235         240
198 Gly Asp Ser Cys Gly Gly Thr Tyr Ser Asn Glu Arg Tyr Ala Gly Val
199          245         250         255
200 Cys Asp Pro Asp Gly Cys Asp Phe Asn Ser Tyr Arg Gln Gly Asn Lys
201          260         265         270
202 Thr Phe Tyr Gly Lys Gly Met Thr Val Asp Thr Thr Lys Lys Ile Thr
203          275         280         285
204 Val Val Thr Gln Phe Leu Lys Asp Ala Asn Gly Asp Leu Gly Glu Ile
205          290         295         300
206 Lys Arg Phe Tyr Val Gln Asp Gly Lys Ile Ile Pro Asn Ser Glu Ser
207 305         310         315         320
208 Thr Ile Pro Gly Val Glu Gly Asn Ser Ile Thr Gln Asp Trp Cys Asp

```

## RAW SEQUENCE LISTING

PATENT APPLICATION: US/10/810,277

DATE: 10/18/2004

TIME: 10:00:39

Input Set : A:\GC794-2-seqlist.txt

Output Set: N:\CRF4\10182004\J810277.raw

```

209          325          330          335
210 Arg Gln Lys Val Ala Phe Gly Asp Ile Asp Asp Phe Asn Arg Lys Gly
211          340          345          350
212 Gly Met Lys Gln Met Gly Lys Ala Leu Ala Gly Pro Met Val Leu Val
213          355          360          365
214 Met Ser Ile Trp Asp Asp His Ala Ser Asn Met Leu Trp Leu Asp Ser
215          370          375          380
216 Thr Phe Pro Val Asp Ala Ala Gly Lys Pro Gly Ala Glu Arg Gly Ala
217 385          390          395          400
218 Cys Pro Thr Thr Ser Gly Val Pro Ala Glu Val Glu Ala Glu Ala Pro
219          405          410          415
220 Asn Ser Asn Val Val Phe Ser Asn Ile Arg Phe Gly Pro Ile Gly Ser
221          420          425          430
222 Thr Val Ala Gly Leu Pro Gly Ala Gly Asn Gly Gly Asn Asn Gly Gly
223          435          440          445
224 Asn Pro Pro Pro Pro Thr Thr Thr Thr Ser Ser Ala Pro Ala Thr Thr
225          450          455          460
226 Thr Thr Ala Ser Ala Gly Pro Lys Ala Gly Arg Trp Gln Gln Cys Gly
227 465          470          475          480
228 Gly Ile Gly Phe Thr Gly Pro Thr Gln Cys Glu Glu Pro Tyr Thr Cys
229          485          490          495
230 Thr Lys Leu Asn Asp Trp Tyr Ser Gln Cys Leu
231          500          505
233 <210> SEQ ID NO: 5
234 <211> LENGTH: 507
235 <212> TYPE: PRT
236 <213> ORGANISM: Humicola grisea
238 <400> SEQUENCE: 5
239 Gln Gln Ala Cys Ser Leu Thr Thr Glu Arg His Pro Ser Leu Ser Trp
240 1          5          10          15
241 Asn Lys Cys Thr Ala Gly Gly Gln Cys Gln Thr Val Gln Ala Ser Ile
242          20          25          30
243 Thr Leu Asp Ser Asn Trp Arg Trp Thr His Gln Val Ser Gly Ser Thr
244          35          40          45
245 Asn Cys Tyr Thr Gly Asn Lys Trp Asp Thr Ser Ile Cys Thr Asp Ala
246          50          55          60
247 Lys Ser Cys Ala Gln Asn Cys Cys Val Asp Gly Ala Asp Tyr Thr Ser
248 65          70          75          80
249 Thr Tyr Gly Ile Thr Thr Asn Gly Asp Ser Leu Ser Leu Lys Phe Val
250          85          90          95
251 Thr Lys Gly Gln His Ser Thr Asn Val Gly Ser Arg Thr Tyr Leu Met
252          100          105          110
253 Asp Gly Glu Asp Lys Tyr Gln Thr Phe Glu Leu Leu Gly Asn Glu Phe
254          115          120          125
255 Thr Phe Asp Val Asp Val Ser Asn Ile Gly Cys Gly Leu Asn Gly Ala
256          130          135          140
257 Leu Tyr Phe Val Ser Met Asp Ala Asp Gly Gly Leu Ser Arg Tyr Pro
258 145          150          155          160
259 Gly Asn Lys Ala Gly Ala Lys Tyr Gly Thr Gly Tyr Cys Asp Ala Gln

```

**VERIFICATION SUMMARY**

PATENT APPLICATION: US/10/810,277

DATE: 10/18/2004

TIME: 10:00:40

Input Set : A:\GC794-2-seqlist.txt

Output Set: N:\CRF4\10182004\J810277.raw